From Volume-based to Value-based Payment: Aspiration Meets Reality

Robert A. Berenson, MD, FACP
Institute Fellow, the Urban Institute
rberenson@urban.org
2015 ABMS National Policy Forum
22 April 2014        Washington, DC
The Presentation Will ----

- Review some definitions and correct some misunderstanding about payment methods and incentives
- Describe the promise and challenges of the broad consensus about the move to paying for value
- Present the *quid pro quo* that Congress is going to require (has already legislated?) in exchange for repeal of the Sustainable Growth Rate (SGR) provision in Medicare
From Volume-based to Value-based Payment

• About the only area in health policy that Republicans and Democrats agree on

• For all of the Republican objections to Obamacare, there generally have not been major Republican objections to the payment/delivery system reform parts of the ACA

• See, for example, the deal in the SGR Repeal Bill contains close to a 20% swing on physician payment based on what is now being called Merit Based Incentive Payment -- the merger of the PQRS, the value-based payment modifier & meaningful use incentives
Some Terminology

• Fee-for-service (FFS) – payments made for each individual service provided during an encounter or hospital stay (actually, individual services that are codified and recognized for payment)
• “FFS Medicare” – the commonly used, if incorrect, term for the traditional Medicare (TM) program to distinguish it from Medicare Advantage. In fact, most payments are not FFS in FFS Medicare
• Volume-based payment – payments that increase as a function of the number of units of services performed – most TM payments
Terminology (cont.)

• Value-based payment – payments that include some level of financial rewards or penalties for measured quality and/or incentives for holding down costs
  – Note that value-based payments are usually placed on top of volume-based payments – not either/or

• Population-based payment – payments made prospectively to a provider responsible for a population of individuals, irrespective of the actual services provided
The Triple Aim

• “The Institute for Healthcare Improvement’s Triple Aim is a framework that describes an approach to optimizing health system performance. It is IHI’s belief that new designs must be developed to simultaneously pursue three dimensions, which we call the “Triple Aim”:
  – Improving the patient experience of care (including quality and satisfaction);
  – Improving the health of populations; and
  – Reducing the per capita cost of health care.”
  – IHI website
Policy Priorities in Medicare (Congressionally directed)

- Value-based purchasing (a euphemism for pay-for-performance) for most provider systems, including physicians
- New payment models in demonstrations: shared savings, bundled episodes, some attention to correcting distortions in the resource-relative value scale-based Medicare Fee Schedule
- New organizational delivery models, esp. accountable care organizations and medical homes
The Prevailing Policy Consensus About Measurement

• After correctly reviewing the important positive effects of quality measurement, finding that “the proof of concept” phase of national quality measurement and public reporting has now been completed, the paper goes on to conclude:

“… As the science has advanced, we now have a surfeit of measures that meet [four rigorous] accountability criteria with which to populate accreditation, public reporting, and pay-for-performance programs.”

Really?

A surfeit -- without any measures of diagnosis accuracy or many about appropriateness or multi-morbidity or issues around toward the end-of-life care (see Atul Gawande, *Being Mortal*)
Berenson View of the Current State of Policy Making

What we measure is considered important and worthy of attention

What we can’t or don’t measure is marginalized or ignored altogether
But What Do We Really Mean by “Value” in Health Care?

• In current health care policy parlance, Value = Quality/Costs and is used in a loose sense to mean a “bigger bang for the buck,”
  – relying on studies showing that the extra 60% the US spends on health services does not buy better health or health care

• But there is no quantitative precision to this equation -- is value increased when quality increases at higher cost?
The Quality Numerator

• Quality is measured differently for different quality items, e.g., % compliance with a process of care standard, 30-day mortality rate for a condition, patient experience from surveys, etc.

• There is no common metric like quality-adjusted life years (QALYS) as used in cost-effectiveness analysis

• As noted, we have good measures in some areas, but few or none in other important clinical domains, e.g., diagnosis errors, appropriateness of services, complex care management

• Identifying a measure gap doesn’t mean it can readily be filled (which Congress doesn’t seem to understand)
The Cost Denominator

- Costs are usually measured as dollars spent but for some purposes can also be measured as the rate of increase in dollars spent, as in “bending the curve”

- Even with something as seemingly straight-forward as dollars spent, there are disagreements on how to measure and report costs, beyond mistaking charges for costs

- Highly unreliable attribution of costs to a clinician or organization, unless you have a population-based metric, such as per person per month spending, for individuals somehow assigned to an organization for care
Dueling aphorisms

• “You can't manage what you can't measure.”
  – Apparently not W. Edwards Deming, to whom this is usually attributed

• “Not everything that can be counted counts, and not everything that counts can be counted.”
  – guess who?
No, Not Albert Einstein but rather a fellow named William Bruce Cameron
“Achieving the Potential of Health Care Performance Measures”

Timely Analysis of Immediate Health Policy Issues, May 2013

By Robert A. Berenson, Peter Pronovost, and Harlan Krumholz

http://www.rwjf.org/content/dam/farm/reports/reports/2013/rwjf406195
Goals of Performance Measurement and Public Reporting

• Enabling patients to make choices about where to seek care
• Providing consumers, purchasers, and taxpayers some level of accountability
• Facilitating quality improvement by providers and providing them motivation to engage
Proliferation of “information brokers” which rate physicians, hospitals and health plans

• But serious questions about validity and reliability of these ratings and rankings

• There was disagreement across four prominent rating systems, with each identifying different sets of high- and low-performing hospitals. The lack of agreement among four national hospital rating systems is likely explained by the fact that each system uses its own rating methods, has a different focus to its ratings, and stresses different measures of performance. – Austin, et al. Health Affairs, March 2015.
Need to distinguish use of measures for public reporting (and P4P) and internal QI

- For internal Quality Improvement, one can be less rigorously accurate - but then organizations should not use the data for promotional purposes, as is common.

- False positive measures can be readily accepted in internal QI – want to cast a wide net and then home in. But false positives can produce reputational harm (and reduced payment) when publicly reported.
Lots of organizations but none with some important measure-related functions

- Advancing the science of performance measurement
- Developing measurement standards
- Setting parameters for accuracy of measures for use in public reporting and P4P
- Coordinating the development of needed measures for various purposes
- NQF comes closest but has no role in measures not submitted to it or in actually developing measures
There are various data sources – all with limitations

- Administrative data, medical records, and patient surveys are the main ones
- Claims data is relatively easy to get and can be inexpensive but lacks needed clinical detail for some quality domains, e.g., diagnosis errors, appropriateness
- Medical record data provides clinical nuance but not standardized and less practical to use – data abstraction very costly
  - EHRs will help but still problems remain
Hospital Value-Based Purchasing

- Ran from 2003-2009 and was based on the Premier Hospital Quality Incentive Demonstration Project – largest hospital P4P program anywhere
- Rewarded performance for AMI, CHF, PN, CABG, and hip and knee replacement
- Focused mostly on processes, e.g., ASA & beta blocker use in AMI
- Bonuses of 1-2% for top 2 deciles – tournament
Premier Demo (cont.)

• In 2007, started also rewarding improvement
• The voluntarily enrolled 261 Premier hospitals initially performed better but not sustained
• Of perhaps greater concern, performance on the process measures used – Medicare core measures – does not predict meaningful differences in outcomes
• Report from NW England suggests modest, positive impact of hospital P4P*

PQRS and the New Value-based Payment Modifier for Physicians

• Currently physicians asked to report 3-4 -- but increasing -- quality measures – reporting starts as bonuses but become penalties for non-reporting in 2015

• About 35% of physicians report as of 2012 – in a program begun in 2007. In addition to reporting burden, it is clear in talking to docs that with a few exceptions, the measures command no respect – most do not reflect on important or valid aspects of many physicians’ core activities

• Policymakers fail to understand that the economics of medical practices and hospitals vary substantially – 1% for hospitals is a big deal; not so for docs, who can generate more FFS revenue
PQRS (cont.)

- The average primary care physician cares for 400 different diagnostic categories in a year; 70 conditions make up 80% of clinical episodes -- yet PQRS relies on a few, mostly screening, prevention and antibiotic usage.
- The measures are becoming more clinically relevant, but raise challenging reliability issues because of case mix bias, self-reporting and small sample sizes.
- Good luck with the cost denominator where costs across the system have to be attributed to a single physician.
The Value-Based Payment Modifier

• Created by ACA, based on PQRS, a P4P approach by Medicare to pay physicians a differential payment based on the quality of care furnished as compared with the cost, producing a global estimate of a physician’s (or medical group’s) value
• Will result in a reward or penalty amounting to 1-2% of payments for groups of 100 physicians or more in 2015 and for all physicians by January 1, 2017
• But in the SGR Repeal bill that, the various PQRS-based and meaningful use based rewards and penalties would be substantially increased in their payment impact
“You don’t like the eating at the ‘Benjamin Arms’, Mrs. Margolies?

“No! It was terrible! Every mouthful was positively poison — and what small portions they give you of it.”

-- Harry Hirshfield, 1927
Behavioral Economists Are Starting to Weigh

• At least with reference to professionals, as opposed to organizations (“corporations are people, my friend”)

• The extrinsic motivation of financial rewards may “crowd out” intrinsic motivation to act in patients’ best interests

• In this context, the concern is with what happens to non-measured quality, especially when – as is the case for many specialties– the PQRS measures are trivial compared to the core of what the specialty is expected to do
Issues in Relying Mostly on Process Measures

- Most available and used measures in “the surfeit of measures” are process measures, with some, important intermediary outcome ones

- Advantages:
  - relatively easy to obtain data
  - *may* have less need for case-mix adjustment
  - based on professional standards and ideally strong evidence, so clinicians should find credible
  - “actionable”
But Many Disadvantages of Reliance on Process Measures

• There are literally thousands of proliferating measures but still few in important quality domains
• Generally, we have measures for preventive care, chronic care management, and some safety items but not effectiveness, coordination, efficiency, and diagnostic accuracy, and patient-important/functional outcomes
• Unfortunately, many available process measures are not necessarily good predictors of outcomes – though some surely are
More Disadvantages of Reliance on Process Measures

- “Teaching to the test”
- High cost of data collection – so inordinate reliance on claims information related to lab and Rx use
- Self-reporting may be inaccurate or overtly gamed
- Need to continually update for changes in evidence – Could freeze innovation where the tail of measurement wags the dog of clinical progress
The Promise and Challenges of True or Intermediate Outcome Measures

• True outcomes are what really matters, especially to patients

• Intermediate outcomes are often easier to attain and can be valid predictors of important outcomes, as with HgbA1C, blood pressure (if done in a standard way and reported accurately), and dialysis measures (Kt/V, Hgb level)

• Focus on improving outcomes permits much broader focus than on specific processes of care

• For example, the literature shows that what matters most in AMI hospital outcomes relates to teamwork, culture, leadership, and other, not strictly clinical, factors

• Promoting “unprofessional” behavior – such as increases in 31-day mortality rates
Specific Challenges in Relying on Outcome Measures

• Risk or case-mix adjustment is crucial – both clinical and social determinants – and difficult
• Data validity, e.g., what is not needed for payment may not be reported
• Surveillance bias – if you look for conditions, you find them. DVTs, La Canicule in France in 2003.
  – And for sure diagnosis errors
• Sample sizes, especially for adverse events
Seven Recommendations for Improving Measurement

• Decisively move from measuring processes to outcomes;
• Use quality measures strategically, adopting other quality improvement; approaches where measures fall short
• Measure quality at the level of the organization, rather than the clinician;
• Measure patient experience with care and patient-reported outcomes as ends in themselves;
• Use measurement to promote the concept of the rapid-learning health care system
Measurement Recommendations (cont.)

- Invest in the “basic science” of measurement development, including an emphasis on anticipating and preventing unintended adverse consequences; and

- Task a single entity with defining standards for measuring and reporting quality and cost data, similar to the role the SEC services for the reporting of corporate financial data, to improve the validity and comparability of publicly-reported quality data.
Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
Background

• **1997:** SGR update formula passed in effort to control volume and “unsustainable” growth in Medicare Part B spending

• **Since 2003:** SGR-mandated physician payment cuts did not go into effect 15 times
  – So far has cost about $150 billion in annual “fixes”
The Bill Would --

• Repeal the SGR, specifying fee updates for 10 years and “improved” payment through a consolidated Merit Based Payment Incentive System (MIPS),
• Set priorities and funding for quality measure development
• Set up an alternative payment system for physicians actively participating with Alternative Payment Methods
• When Congress could not pass this legislation last year certain other elements that are relevant to the existing Medicare Physician Fee Schedule were pulled out and put into the SGR “patch” bill that provided a “doc fix” till March 2015 – including a GAO study of the RUC, establishing a form of prior authorization for imaging targeted to outliers
“Stabilizing” Fee Updates

• Repeal SGR, and 21% fee cut retroactive to April 1, 2015
• 7/2015- 12/2019: Annual fee update 0.5%
• 2020-2025: 0%
  – Payment increases take place through MIPS
• ≥ 2026: 0.25% update (non-qualifying APM)
  – 0.75 % if APM participation
• MedPAC studies in 2017 and 2021 that assess how spending in Part B relates to total A, B, and D spending
Unified Merit Based Incentive Payment System (MIPS)

• Combines the 3 current incentive programs:
  – Physician Quality Reporting System (PQRS) – quality
  – Value-Based Modifier (VBM) – quality & resource use
  – Meaningful Use (EHR)

• Applies to payments after January 1, 2018 – current programs in use till then and sun-setted
Unified Merit Based Incentive Payment System (MIPS)

- Applies to all the types of health professionals receiving fee schedule $’s
- Excludes those with too few Medicare patients and those who have threshold level of payment through alternative payment methods (APMs)
- May participate through EHR use, qualified clinical data registries and/or through group, “virtual” group or affiliation with a facility or hospital
MIPS Assessment Categories

• Quality (30%)
  – Current measures
  – Solicitation of new measures
  – Qualified clinical data registries

• Resource Use (30%)
  – Current VBM measures
  – Develop new measures
  – Link cost of services to a professional: Allow for reporting of role in treatment & type of treatment
  – Research on risk adjustment
MIPS Assessment Categories (cont.)

• Meaningful Use (MU) (25%, although some variation)
  – Current system use
  – Reporting through certified EHR systems for MIPS are deemed to meet MU component

• Clinical Practice Improvement Activities (15%)
  – Credit for engaging in clinical practice improvement activities (expanded practice areas, population management, care coordination, beneficiary engagement, patient safety)
  – Activities must be applicable to all specialties & attainable for small practices and underserved areas
  – Credit if already doing
  – Encourages activities that facilitate future APM participation
The term ‘clinical practice improvement activity’ means “an activity that relevant eligible professional organizations and other relevant stakeholders identify as improving clinical practice or care delivery and that the Secretary determines, when effectively executed, is likely to result in improved outcomes”.

The term ‘eligible professional organization’ means “a professional organization as defined by nationally recognized specialty boards of certification or equivalent certification boards”.
ANNUAL LIST OF QUALITY MEASURES AVAILABLE FOR MIPS ASSESSMENT

• Call for quality measures:
  – Eligible professional organizations and other relevant stakeholders shall be requested to identify and submit quality measures to be considered for selection in the annual list of quality measures and to identify and submit updates to the measures on such list.
  – Measures may be submitted regardless of whether such measures were previously published in a proposed rule or endorsed by an entity with a contract under section 1890(a).
Composite Performance Score

• 1-100 score based on four assessment categories
• Weights adjusted to providers ability to report on particular measures
• Credit for annual improvement somehow to be factored in (quality, resource use, +/- clinical practice improvement)
Payment Adjustment

- Based on calculated performance threshold
  - Mean or median of all MIPS eligible professionals composite performance scores
  - Calculated prior to start of each performance period
- Linear distribution based on comparison to performance threshold – i.e., EPs whose composite performance scores fall above the performance threshold (mean or median of composite) will get + positive adjustments and the converse
- Entire funding pool is paid out each year
MIPS Payment Adjustment

• Negative adjustments capped
  – Those at 0-25% of threshold get maximum negative adjustment
    • 2019: 4%
    • 2020: 5%
    • 2021: 7%
    • 2022: 9%

• Positive
  – Maximum: 3 X annual cap for negative adjustment
  – Additional payment if 25% above performance threshold
    • Capped at $500 million / year (2019-2023)
• GAO evaluations of MIPS with reports in 2019 and 2021– incl. variations by practice size, geography, patient mix
APM Participation

• 5% bonus (2019-2023) if physician has significant APM participation
  – “ if x percent of payments for covered professional services were
    attributable to such services… through an entity that participates in
    an eligible APM with respect to such services.” [emphasis added]
  – APM must involve risk of financial losses & have a quality
    measure component
  – Part of a PCMH exempt from the downside risk if CMMI finds it
    works in Medicare
  – Two tracks – % of Medicare revenue and % of all/other payer
    revenue with specs related to presence or not of M & M APM
    opportunities

• Exempt EPs are excluded from MIPS & most EHR requirements
• Special emphasis on testing APMs with specialists & small practices
  and that align with private and state-based payer initiatives
HHS Framework for the Evolution of Payment Models

- Category 1—fee-for-service with no link of payment to quality
- Category 2—fee-for-service with a link of payment to quality
- Category 3—alternative payment models built on fee-for-service architecture
- Category 4—population-based payment

“Value-based purchasing includes payments made in categories 2 through 4.”
HHS Jan 26 “Historic Announcement” of Goals and Timeline for Value Payments

- 30% of traditional Medicare payments tied to value thru APMs (categories 3,4) by the end of 2016, and 50% by 2018
- 85% tied to value (categories 2-4) by 2016 and 90% by 2018
- CMS says “the majority of Medicare payments now are linked to quality” – that is true only by crediting any use of P4P in a payment system as a link to quality -- even something trivial affecting few services